

# Using Correlated Subset Structure for Compressive Sensing Recovery

Atul Divekar

Alcatel-Lucent

Naperville, IL 60563

Email: atul.divekar@alcatel-lucent.com

Deanna Needell

Department of Mathematics

Claremont McKenna College

Claremont, CA 91711

Email: dneedell@cmc.edu

**Abstract**—Compressive sensing is a methodology for the reconstruction of sparse or compressible signals using far fewer samples than required by the Nyquist criterion. However, many of the results in compressive sensing concern random sampling matrices such as Gaussian and Bernoulli matrices. In common physically feasible signal acquisition and reconstruction scenarios such as super-resolution of images, the sensing matrix has a non-random structure with highly correlated columns. Here we present a compressive sensing recovery algorithm that exploits this correlation structure. We provide algorithmic justification as well as empirical comparisons.

## I. INTRODUCTION

Consider the problem of image super-resolution, where one or more low-resolution images of a scene are used to synthesize a single image of higher resolution. If multiple images are used, they are commonly assumed to be subpixel-shifted and downsampled versions of the original high resolution image that is to be reconstructed [1]. Alternatively, super-resolution from a single low resolution image using a dictionary of image patches and compressive sensing recovery has been proposed in [2]. The relationship between the available low resolution and desired high resolution image is commonly modeled by a linear filtering and downsampling operation. Suppose that we wish to reconstruct a size  $N \times N$  high resolution image from a lower resolution image, for example of size  $\frac{N}{2} \times \frac{N}{2}$ , or smaller. Let  $x$  and  $y$  represent the vectorized high and low resolution images respectively. We model the formation of  $y$  from  $x$  by the equation  $y = SHx + \eta$  where  $\eta$  is the sensor noise,  $S$  is a downsampling matrix of size  $\frac{N}{2} \times \frac{N}{2}$  by  $N^2$ , and  $H$  is a  $N^2$  by  $N^2$  matrix that represents the filtering (antialiasing) operation. In order to consider super-resolution as a compressive sensing recovery problem we write  $x = \Psi c$  where  $\Psi$  is a sparsifying basis for the class of images under consideration and  $c$  is the coefficient vector corresponding to image  $x$  with respect to the basis  $\Psi$ . In the simplest case,  $\Psi$  is an  $N^2 \times N^2$  orthogonal matrix, but can also be generalized to an overcomplete dictionary.

We have

$$y = SH\Psi c + \eta = \Phi c + \eta,$$

where  $\Phi = SH\Psi$  is the sampling matrix.

Most of the work in the compressive sensing literature assumes  $\Phi$  to be random matrix, such as a partial DFT or one

drawn from a Gaussian or Bernoulli distribution. However, in this scenario the matrix is not random, but instead has correlated columns whose structure we wish to exploit to improve compressive sensing recovery. Here we assume that  $H$  is not a perfect low pass filter, so that it is possible for  $\Phi = SH\Psi$  to preserve enough high frequency information for recovery to be possible;  $SH$  and  $\Psi$  have sufficient incoherency to allow  $c$  to be recovered with acceptable error.

Compressed sensing provides techniques for stable sparse recovery [3]–[5], but results for coherent sensing matrices have been limited [6], [7].

**Organization.** The structure we wish to exploit is first described. Then we present algorithms that take advantage of this structure for compressive sensing recovery.

## II. CORRELATION STRUCTURE

Typical examples of sparsifying bases  $\Psi$  for images are wavelets and blockwise discrete cosine transform bases. Images exhibit correlation at each scale: neighboring pixels are heavily correlated except across edges, local averages of neighboring blocks are heavily correlated except across edges, and so on. This makes wavelet-like bases, which have locally restricted atoms, suitable for sparsifying the image. For the super-resolution setting with the low resolution image of size  $\frac{N}{2} \times \frac{N}{2}$ , the rows of  $SH$  consist of shifted versions of the filtering kernel with shifts of 2 horizontally and vertically. Due to the localized nature of wavelet bases, we expect columns of  $\Phi$  that correspond to spatially distant bases in  $\Psi$  to have little correlation. If  $\Psi$  is a tree structured orthogonal wavelet basis matrix, columns of  $\Psi$  that overlap spatially are orthogonal, however when filtered by  $H$ , they result in significant correlation. Then we expect columns in  $\Phi$  to show significant correlation in tree structured patterns.

We illustrate this with an example. For simplicity we consider only one-dimensional signals, though the discussion is equally valid for images. Suppose that  $\Psi$  is a  $256 \times 256$  matrix whose columns consist of the length 256 Haar basis vectors, and  $SH$  is a  $128 \times 256$  matrix obtained by shifting the filter kernel  $h = \{0.1, 0.2, 0.4, 0.2, 0.1\}$  by two from one row to the next.  $SH$  represents the filtering and downsampling operation that generates the low resolution signal  $y = SHx$  from the length 256 signal  $x$ . Then  $\Phi = SH\Psi$  is the sampling matrix.

Fig. 1 shows the absolute values of the correlation matrix  $C = \Phi^* \Phi$  (here and throughout  $A^*$  denotes the adjoint of  $A$ ). This shows that only a small number of pairs of columns of  $\Phi$  are strongly correlated to each other. Each filtered wavelet basis is correlated with other spatially overlapping bases at coarser and finer scale and in the immediate neighborhood, but has no correlation with spatially distant bases.

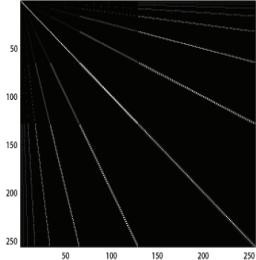


Fig. 1: Absolute values of  $\Phi^* \Phi$ .

More generally, consider compressive sensing recovery where the columns of the sampling matrix  $\Phi$  can be grouped into nearly-isolated sets, such that correlation among pairs of columns within a set may be significant, but correlation between two columns that belong to different sets is relatively small. How does one exploit this structure to efficiently reconstruct the signal?

One of the central results in compressive sensing is that if matrix  $\Phi$  exhibits a property called the Restricted Isometry Property (RIP) [8]–[10], convex optimization can recover the sparse signal exactly [11], [12] via

$$\min \|c\|_1 \text{ such that } y = \Phi c. \quad (1)$$

However, the sampling matrix  $\Phi = H\Psi$  described above does not obey the RIP and these results are not readily applicable. On the other hand, it is commonly found in practical applications and has a structure that could be exploited.

Before considering the above problem, a simple modification to CoSaMP [13] is presented that provides some improvement in recovery performance. This algorithm, called Partial Inversion (PartInv) and described by Algorithm 1, also indicates how the above described structure could be exploited.

### III. PARTIAL INVERSION

Consider the usual CS setting: Given a length  $M$  sample vector  $y = \Phi c + \eta$  where  $\Phi$  is an  $M \times N$  sampling matrix and  $c$  a length  $N$  vector with sparsity  $K < M$ , we wish to obtain the best  $K$ -sparse approximation  $\hat{c}$  to  $c$ . At each step let  $I$  be an index set, so that for example,  $\hat{c}_I$  represents an estimate of the components of  $c$  corresponding to the column indices in  $I$ .  $\hat{c}$  by itself is an estimate for all the columns  $\{1..N\}$ . Let  $L$  for  $K \leq L < M$  be an adjustable parameter for the size of the set  $I$ . We get good results with  $L = \max\{K, 0.8M\}$ . Let  $\Phi_I$  denote the matrix of columns from  $\Phi$  corresponding to indices in set the  $I$ . Let  $\tilde{I} = \{1..N\} \setminus I$  denote the complement of  $I$ . For any full rank matrix  $A$ , define  $A^\dagger = (A^* A)^{-1} A^*$ .

---

**Algorithm 1** Given  $y = \Phi c$ , return best  $K$ -sparse approximation  $\hat{c}$

---

- 1:  $\hat{c} \leftarrow \Phi^* y; I^0 \leftarrow$  indices of the  $L$ -largest magnitudes of  $\hat{c}; k \leftarrow 0$
  - 2: **while** Stopping condition not met **do**
  - 3:    $\hat{c}_{I^{(k)}} \leftarrow \Phi_{I^{(k)}}^\dagger y$
  - 4:    $r \leftarrow y - \Phi_{I^{(k)}} \hat{c}_{I^{(k)}}$
  - 5:    $J^{(k)} \leftarrow I^{(k)}$
  - 6:    $P \leftarrow [I - \Phi_{I^{(k)}} \Phi_{I^{(k)}}^\dagger] \Phi_{J^{(k)}}$
  - 7:    $\hat{c}_{J^{(k)}} \leftarrow P^* r$
  - 8:    $I^{(k+1)} \leftarrow$  indices of the  $L$ -largest magnitude components of  $\hat{c}$ .
  - 9:    $k \leftarrow k + 1$
  - 10: **end while**
- 

For the noiseless case  $\eta = 0$ , the stopping condition can be obtained by testing the magnitude of  $r_2 = y - \Phi \hat{c}$  at the start of each iteration. If set  $I$  does not vary from one iteration to the next, the algorithm cannot progress further and can be stopped immediately. In practice the inversion of line 3 can be done efficiently by Richardson's algorithm (see e.g. Sec. 7.2 of [14]). If  $|J^{(k)}| < M$ , we may replace line 6 by  $\hat{c}_{J^{(k)}} \leftarrow \Phi_{J^{(k)}}^\dagger r$ .

This algorithm demonstrates improvement relative to CoSaMP when the accurate recovery region is considered on a plot of  $\frac{K}{M}$  versus  $\frac{M}{N}$ . The motivation is the following (for simplicity we drop the iteration indicator  $k$ ): From line 3,

$$\hat{c}_I = \Phi_I^\dagger y \quad (2)$$

$$= c_I + (\Phi_I^* \Phi_I)^{-1} \Phi_I^* \Phi_{\tilde{I}} c_{\tilde{I}}. \quad (3)$$

Compare this to the estimator  $\hat{c}_I = \Phi_I^* r$  used in CoSaMP. When  $r = y$ , we have

$$\hat{c}_I = \Phi_I^* y \quad (4)$$

$$= \Phi_I^* \Phi_I c_I + \Phi_I^* \Phi_{\tilde{I}} c_{\tilde{I}} \quad (5)$$

$$= c_I + (\Phi_I^* \Phi_I - I)c_I + \Phi_I^* \Phi_{\tilde{I}} c_{\tilde{I}} \quad (6)$$

If the index set  $I$  contains several nonzero coefficients (which we hope is true), then  $(\Phi_I^* \Phi_I - I)c_I$ , which results from the mutual interference between the columns of  $\Phi_I$ , is significant and is a source of noise in  $\hat{c}_I$ . This term is eliminated in (2). Partial inversion does add  $(\Phi_I^* \Phi_I)^{-1}$  to the remaining noise term, however, the singular values of this term can be kept from significantly amplifying the noise term by a conservative choice of  $L$ , the size of the index set  $I$ . The improved estimate  $\hat{c}_I$  further produces an improved estimate  $\hat{c}_{J^{(k)}}$ , which leads to a better selection of nonzero coefficients in the next iteration.

The expression (2) also indicates how the correlation structure may be used to improve recovery. The noise term  $(\Phi_I^* \Phi_I)^{-1} \Phi_I^* \Phi_{\tilde{I}} c_{\tilde{I}}$  depends upon the correlation between the sets  $\Phi_I$  and  $\Phi_{\tilde{I}}$  given by  $\Phi_I^* \Phi_{\tilde{I}}$ . This correlation is weak if  $\Phi_I$  and  $\Phi_{\tilde{I}}$  are sufficiently spread. However, the correlation is

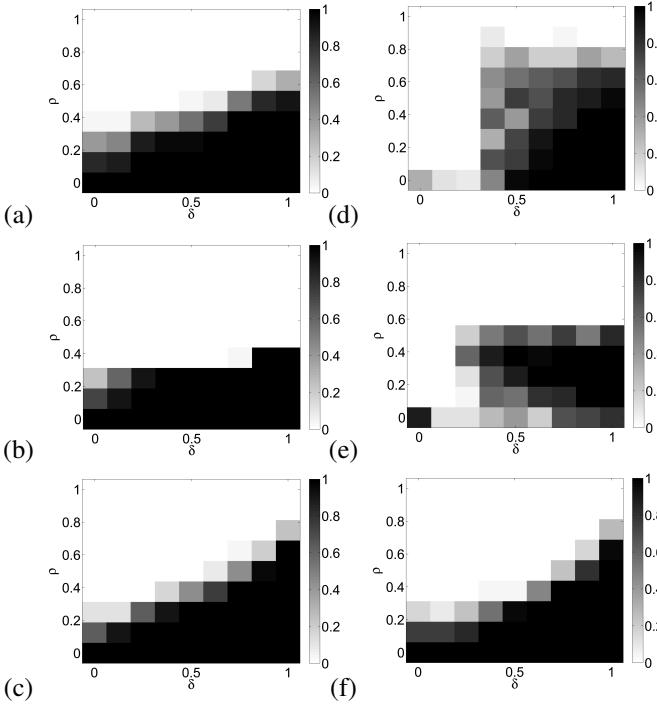


Fig. 2: Proportion of successes on Gaussian matrices using (a) PartInv, (b) CoSaMP and (c)  $\ell_1$ -minimization, and proportion of successes on correlated column subset matrices using (d) PartInv, (e) CoSaMP and (f)  $\ell_1$ -minimization for various values of  $\delta = \frac{M}{N} \in (0, 1)$  (horizontal axis) and  $\rho = \frac{K}{M} \in (0, 1)$  (vertical axis).

likely to remain large if  $L$  is significant compared to  $M$ , as will be the case when  $\frac{K}{M}$  is large.

#### IV. EXPERIMENTAL COMPARISON

We compare the recovery performance of Partial Inversion with CoSaMP and convex optimization (1) for two classes of matrices: Gaussian random matrices, and matrices constructed to have highly correlated subsets of columns with low correlation across subsets.

In the first case, we construct  $M$  by  $N$  matrices with  $N(0, 1)$  elements along with the coefficient vector  $c$  containing  $K$  nonzero entries taken from a  $N(0, 1)$  distribution. The nonzero locations are selected uniformly at random from  $\{1\dots N\}$ . Each column in each matrix is normalized to have unit  $l_2$  norm. We set  $N = 256$  and vary  $\delta = \frac{M}{N}$  from 0.1 to 0.9 in steps of 0.1. For each  $\delta$  we vary  $\rho = \frac{K}{M}$  from 0.1 to 0.9 in steps of 0.1. For each  $(\delta, \rho)$  point we carry out 25 trials, and declare success if  $\frac{1}{N}||c - \hat{c}||^2 < 10^{-5}$ . For PartInv we considered two cases for the size of subset  $I$ :  $L = S$  and  $L = \max\{S, 0.8M\}$ . We see better performance in the  $L = S$  case. For  $\ell_1$  minimization we use the  $\ell_1$ -magic package [15]. We show the results in Fig. 2.

In the second case, we construct  $M$  by  $N$  matrices with  $N = 256$  and variable  $M$  and a block diagonal structure. The columns are divided into 16 column subsets. In each subset we set  $M/16$  rows to 1. In addition, to every element of the matrix we add noise drawn from a zero-mean normal

distribution with variance 0.0025. This produces heavy intra-subset correlation and light correlation across subsets. We let the coefficient vector  $c$  contain  $S$  nonzeros elements drawn from a  $N(0, 1)$  distribution. We select 4 of the 16 subsets at random and in each subset select  $\frac{S}{4}$  of the indices to have nonzero values, again uniformly at random. If some of the nonzeros were left over, they are accommodated in a fifth subset. For PartInv we set  $L = \max\{S, 0.8M\}$ . The results are also depicted in Fig. 2.

#### V. RECOVERY OF COEFFICIENTS CONCENTRATED ON WAVELET TREES

We next use Partial Inversion to recover nonzero coefficients that are concentrated on wavelet trees, which is commonly seen when a signal or image with discontinuities is decomposed in a wavelet basis. When the coefficients are concentrated on an isolated set (a set of columns that have low correlation with columns outside the set), a setwise estimator is especially useful to identify the sets on which the coefficients are nonzero. Consider the 2D wavelet case. Suppose that  $I$  is the index set of columns of the wavelet basis belonging to a particular tree rooted at a coarse scale and containing finer scale coefficients. We have

$$z_I = \Phi_I^*y = \Phi_I^*\Phi_{ICI} + \Phi_I^*\Phi_{\bar{I}}c_{\bar{I}}. \quad (7)$$

Because  $\Phi_I$  is relatively isolated from the columns in  $\Phi_{\bar{I}}$ , the second term is small, and because most of the elements of  $c_I$  are nonzero, the first term is large. This is further intensified by the mutual correlation of the columns of  $\Phi_I$  which is high because of the spatial overlap of the support of the wavelet bases in the tree. This motivates a simple selection criterion for measuring the strength of the nonzero coefficients in each wavelet tree  $I$ :  $s_I = \sum_{j \in I} |z_j|$ . We use this criterion along with PartInv to select wavelet trees that are known to be nonzero. We denote the number of subsets by SETNUM.

We modify the PartInv algorithm to use this estimator.

#### VI. EXPERIMENTAL RESULTS

To test this algorithm, we use the Daubechies-5 wavelet basis in two dimensions over  $32 \times 32$  size patches with 5 levels of decomposition. This gives a size 1024 by 1024 matrix  $\Psi$ . We divide this matrix into 49 sets: 1 set of the coarsest scale coefficients in a block of size  $4 \times 4$  containing the two coarsest scales, and 48 other sets rooted at the coefficients at the next finer scale. Each of these sets contains 21 ( $1+4+16$ ) coefficients in a quadtree structure. To create matrix  $\Phi$  we first apply a blurring filter  $H$  with a symmetric  $3 \times 3$  kernel that is close to a delta function. This simulates practical optical sampling acquisition effects such as diffraction and helps prevent rank deficiency problems when carrying out inversion. We use different 2D sampling patterns to carry out the subsampling operation represented by matrix  $S$ . Hence the acquisition process is represented by  $y = \Phi c$  where  $\Phi = SH\Psi$ . The sampling patterns are shown in Table I for each sampling rate  $\delta = \frac{M}{N}$  used to generate the results.

**Algorithm 2** Given  $y = \Phi c$ , with  $K$  nonzero coefficients concentrated on wavelet trees, return best  $K$ -sparse approximation  $\hat{c}$

```

1:  $\hat{c} \leftarrow \Phi^* y$ ;
2:  $k \leftarrow -1$ 
3: for  $j = 1 \rightarrow \text{SETNUM}$  do
4:    $s_j \leftarrow \sum_{l \in I_j} |\hat{c}_l|$ 
5: end for
6:  $I^{k+1} \leftarrow$  indices of columns contained in the sets with the
   largest magnitude  $s_k$ , to include at least  $K$  coefficients.
7:  $k \leftarrow k + 1$ 
8: while Stopping condition not met do
9:    $\hat{c}_{I^{(k)}} \leftarrow \Phi_{I^{(k)}}^\dagger y$ 
10:   $r \leftarrow y - \Phi_{I^{(k)}} \hat{c}_{I^{(k)}}$ 
11:   $J^{(k)} \leftarrow I^{(k)}$ 
12:   $P \leftarrow [Id - \Phi_{I^{(k)}} \Phi_{I^{(k)}}^\dagger] \Phi_{J^{(k)}}$ 
13:   $\hat{c}_{J^{(k)}} \leftarrow P^* r$ 
14:  Repeat lines 3 – 6
15:   $k \leftarrow k + 1$ 
16: end while

```

Each pattern is replicated 8 times in horizontal and vertical directions to give the  $32 \times 32$  sampling pattern used for matrix  $S$ . The filter kernel is a  $3 \times 3$  kernel with 0.29 at the center and 0.02 in other locations.

0	0	0	0
0	1	0	0
0	0	0	0
0	0	0	1

$$(a) \delta = \frac{2}{16}$$

1	0	0	0
0	0	1	0
0	1	0	0
0	0	0	1

$$(b) \delta = \frac{4}{16}$$

1	0	1	0
0	1	0	1
1	0	1	0
0	0	1	0

$$(c) \delta = \frac{6}{16}$$

1	0	1	0
0	1	0	1
1	0	1	0
0	1	0	1

$$(d) \delta = \frac{8}{16}$$

0	1	0	1
1	0	1	0
0	1	1	1
1	1	0	1

$$(e) \delta = \frac{10}{16}$$

1	1	0	1
0	1	1	1
1	1	0	1
0	1	1	1

$$(f) \delta = \frac{12}{16}$$

1	1	1	1
1	1	0	1
1	1	1	1
0	1	1	1

$$(g) \delta = \frac{14}{16}$$

TABLE I: Sampling Patterns

The results are shown in Fig. 3. For each data point we carry out 100 trials. We declare success if  $\frac{1}{N} \|c - \hat{c}\|^2 < 10^{-5}$  where  $N = 32 \times 32$ . This shows improvement in selection performance with the sum estimator. This estimator is much simpler than the Condensing Sort and Select Algorithm (CSSA) used in [16] to incorporate wavelet tree structure in CoSAMP but gives better results.

## VII. CONCLUSION

We consider methods of compressive sensing recovery for sampling matrices that have subsets of columns that are

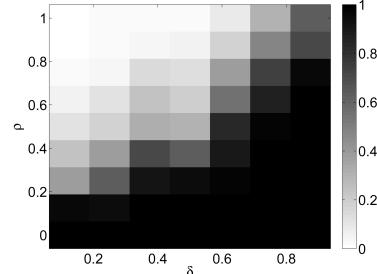


Fig. 3: Proportion of successes out of 100 trials with nonzero coefficients concentrated on wavelet trees.

strongly intra-correlated, but show low correlation with other subsets. This structure commonly arises in physical sample acquisition/reconstruction scenarios such as image super-resolution. We describe Partial Inversion, an algorithm that improves compressive sensing recovery by removing a source of noise in the initial estimator, and demonstrate its performance by simulations on Gaussian and correlated column subset matrices. We consider compressive sensing recovery when the nonzero coefficients are concentrated on wavelet trees, and demonstrate a simple estimator that improves selection of the trees that carry the nonzero coefficients.

## REFERENCES

- [1] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, “Advances and challenges in super-resolution,” *Int. J. Imag. Syst. Tech.*, vol. 14, no. 2, pp. 47–57, 2004.
- [2] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image superresolution via sparse representation,” *IEEE T. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [3] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE T. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [4] E. Candes, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [5] D. Donoho and P. Stark, “Uncertainty principles and signal recovery,” *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, 1989.
- [6] E. Candes, Y. Eldar, D. Needell, and P. Randall, “Compressed sensing with coherent and redundant dictionaries,” *Appl. Comput. Harmon. A.*, vol. 31, no. 1, pp. 59–73, 2011.
- [7] E. Candes and C. Fernandez-Granda, “Towards a mathematical theory of super-resolution,” *Preprint*, 2012.
- [8] E. Candes and T. Tao, “Decoding by Linear Programming,” *IEEE T. Inform. Theory*, vol. 51, no. 12, pp. 4203 – 4215, dec. 2005.
- [9] M. Rudelson and R. Vershynin, “On sparse reconstruction from Fourier and Gaussian measurements,” *Comm. Pure Appl. Math.*, vol. 61, pp. 1025–1045, 2008.
- [10] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constr. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.
- [11] E. Candes and T. Tao, “Near optimal signal recovery from random projections: Universal encoding strategies?” *IEEE T. Inform. Theory*, vol. 52, pp. 5406–5425, Dec. 2006.
- [12] E. Candes, “The Restricted Isometry Property and its implications for Compressed Sensing,” *Cr. Acad. Sci. I-Math.*, pp. 589–592, Dec. 2008.
- [13] D. Needell and J. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Appl. Comput. Harmon. A.*, vol. 26, pp. 301–321, 2009.
- [14] Å. Björck, *Numerical Methods for Least Squares Problems*. Philadelphia: SIAM, 1996.
- [15]  $\ell_1$ -magic. [Online]. Available: <http://www.acm.caltech.edu/l1magic/>
- [16] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, “Model-based compressive sensing,” *IEEE T. Inform. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.